Locality and Mentality in Everett Interpretations I: Albert and Loewer's Many Minds

Laura Felline, Guido Bacciagaluppi

Locality and Mentality in Everett Interpretations I: Albert and Loewer's Many Minds

Laura Felline* and Guido Bacciagaluppi†

2 September 2013

Abstract

This is the first of two papers reviewing and analysing the approach to locality and to mind-body dualism proposed in Everett interpretations of quantum mechanics. The planned companion paper will focus on the contemporary decoherence-based approaches to Everett. This paper instead treats the explicitly mentalistic Many Minds Interpretation proposed by David Albert and Barry Loewer (Albert and Loewer 1988). In particular, we investigate what kind of supervenience of the mind on the body is implied by Albert and Loewer's Many Minds Interpretation, and how the interpretation of the related 'mindless hulks' problem affects the issue of locality within this interpretation.

1 Introduction

There are many issues whose treatment in Everett interpretations of quantum mechanics is distinctive. Two such issues are locality and mentality. Everett interpretations are generally thought to be local, either in the sense that Bell's theorem does not apply to them (because there are no unique outcomes in Everett), or because — since Everett interpretations do not modify

*Institut Supérieur de Philosophie, Université Catholique de Louvain. Address for correspondence: ISP, Place du Cardinal Mercier 14, B-1348 Louvain-la-Neuve, Belgium (email: laura.felline@uclouvain.be).

†Department of Philosophy, University of Aberdeen; Institut d'Histoire et de Philosophie des Sciences et des Techniques (CNRS, Paris 1, ENS); and Collegio Dottorale in Filosofia, Epistemologia e Storia della Cultura, Università di Cagliari. Address for correspondence: Department of Philosophy, University of Aberdeen, Old Brewery, High Street, Aberdeen AB24 3UB, Scotland, U.K. (e-mail: g.bacciagaluppi@abdn.ac.uk).

quantum theory — if the quantum theory under consideration is Lorentz-invariant, then so is the interpretation. Plausible as these intuitions may be, there is surprisingly little literature devoted to analysing this question in detail (exceptions are Bacciagaluppi 2002 and Wallace 2012, Chap. 8). Similarly, of all 'realist' interpretations of quantum theory, Everett interpretations are the ones in which the relation between the mental and the physical is most often discussed, and ideas in the philosophy of mind are most often brought in (e.g. Wallace 2003).[1] Furthermore, the two issues are often linked, especially in so-called 'many minds' variants of Everett. For instance, Albert and Loewer (1988) motivate their own version of the many minds view from desiderata about mind-body dualism. And Zeh — one of the pioneers of the theory of decoherence — takes his own version of the many minds view as the most natural extension of standard views on psycho-physical parallelism when one reduces the decohering wavefunction of the universe to the local density operator of observers' brains (Zeh 2000).

This paper (together with a planned companion) wishes to contribute towards clarifying the questions of locality and mentality in Everett, and in particular of their interplay.

Specifically, this paper will focus on the many minds interpretation (MMI) put forward by David Albert and Barry Loewer (1988), for which the issue of mind-body dualism and supervenience was an explicit motivation. We shall begin in section 2 by reviewing Albert and Loewer's motivations (with particular reference to the so-called 'mindless hulks' problem), and how as a consequence they develop their interpretation. In section 3 we shall then describe in what sense the MMI is local. (Sections 2 and 3 largely rehearse standard material; see also Albert (1992) and Barrett (1999).) In section 4 we return to the 'mindless hulks' problem, and develop an analysis of how various ways of understanding (or misunderstanding) this problem

[1] Notice, however, that the intersections between quantum theory and the mental vs physical literature are far from being restricted to the Everettian context. On the contrary, they include both attempts to use quantum theory to better understand consciousness (see Atmanspacher 2011 for a review of such attempts) and, most relevantly for the topic of this paper, attempts to solve issues in the foundations of quantum theory by elaborating on mind-body dualism and supervenience. Probably the most discussed and famous example of the latter was Eugene Wigner's (1961) suggestion for a solution to the measurement problem. According to such a proposal, the collapse of the wave function is caused by the mind of the observer apprehending the result of a measure on a quantum system. Note in particular the strong mind-body dualism of this proposal: if the observer's consciousness causes the system+apparatus+brain wave function to collapse into a determinate result state, then the observer's belief does not supervene on the physical state of her brain (see Barrett 2006).

bear on the locality of the corresponding variants of the MMI. Finally, section 5 discusses the notion of 'weak nonlocality' introduced by Hemmo and Pitowsky (2003) in their discussion of the MMI. In the companion to this paper, we plan to discuss the explicitly decoherence-based approaches of Saunders (1995), Wallace (2002; 2003) and others ('many worlds'), and of Zeh (2000) ('many minds'), in particular the local and global aspects of the role played by decoherence in these approaches.

2 Albert and Loewer's Many Minds Interpretation

2.1 Making sense of probabilities in Everett

Albert and Loewer's purpose in developing the MMI was to present an interpretation of the Relative State Theory (Everett 1957; 1973) that could provide a valid alternative to the then fairly widespread view that worlds literally split, originally put forward by DeWitt and Graham (DeWitt 1971; DeWitt and Graham 1973). The latter was considered unacceptable by Albert and Loewer due to three fundamental problems, which they term the 'democracy of bases' problem, the 'conservation of mass' problem and the 'determinism' problem.

The first of these problems is that, in the splitting worlds interpretation, the splitting of the original world seems to privilege a particular basis (the one corresponding to the measured observable), while the quantum mechanical formalism does not include any such privileged basis.

The 'conservation of mass' problem concerns the fact that, according to Schrödinger's equation, the mass-energy of the combined observed system and measurement apparatus is the same before and after the measurement, while in the splitting worlds interpretation the total mass-energy would seem to increase with each splitting (with every measurement process).

The 'determinism' problem calls into play the interpretation of probabilities. It is illustrated by Albert and Loewer in the following terms: 'since, according to the [splitting worlds interpretation], it is certain that all outcomes of the measurement will occur and will be observed by successors of [the observer], what can be meant by saying that the probability of a particular outcome $= c^2$?' (1988, p. 201). Albert and Loewer's development of the MMI was driven by an attempt to solve these three problems, and further the ‘mindless hulks’ problem, which derives directly from the last of the three.

According to Albert and Loewer it is impossible to make sense of probabilities in Everettian quantum mechanics by relying only on deterministic dynamical equations. A first problem is that the evolution of the wave function does not provide us with a rule for the transtemporal identity of branches. In DeWitt and Graham’s interpretation, for instance, the only thing we can say about worlds is that, at a given instant t, to each component of the universal wave function corresponds a world in that state, but nothing enables us to say that the ‘worlds’ existing at t are the same worlds that exist at a later instant t'. On the other hand, without transtemporal identity of branches there is no hope of making sense of statements like ‘the probability that I will register spin-up is $= c^2$’, for nothing allows us to identify *me* before the measurement with any *me* existing after the measurement.

Moreover, given the deterministic character of the dynamical equations, according to Albert and Loewer the addition of the transtemporal identity of branches is a necessary but not sufficient condition for a meaningful interpretation of probabilistic statements. This is the ‘determinism problem’ mentioned already: if all the outcomes of a measurement will occur with certainty, the probability for each outcome should be equal to 1, not to some c^2. According to Albert and Loewer, ‘if probability is to be introduced into the picture, it must necessarily be by *adding* something to the interpretation’ (p. 201). However, to ‘add something to the interpretation’ typically means to forsake the central Everettian idea that the wave function is a complete description of the physical world. Thus the problem arises of how to make sense of probability and, at the same time, maintain intact the Everettian postulate of the completeness of the wave function.

Keeping Albert and Loewer’s view of the determinism problem in mind, we shall now see their proposal for its solution. They begin with an intermediate construction, that of the ‘single mind’ interpretation.

2.2 The single mind interpretation

The construction of this view starts with two basic postulates:

1. The universal wave function provides a complete physical description of reality.

2. Through introspection we are able to obtain reliable data regarding our beliefs. Since introspection suggests that we always have well-defined beliefs, we infer that these cannot enter a superposed state.

According to postulate 1, our bodies are generally in a superposition of different brain states,[2] but according to postulate 2, our mind is never in a superposition of the corresponding belief states. This, in turn, implies that the following assumption fails:

M. 'The state wherein A believes that spin-x = up and the state wherein A believes that spin-x = down are identical with certain physical states of A's brain.'

That is, according to Albert and Loewer, the desired interpretation of quantum theory has to be a *dualist* theory of mind and body.

The first proposal Albert and Loewer advance is the *single mind interpretation*.[3] Its basic postulates are:

a. The universal wave function provides a complete description of physical reality.

b. Every sentient physical system is associated with a nonphysical entity called *mind*, which is never in a superposition of belief states. Our state of consciousness corresponds to the state of our mind.

c. The evolution of the mind during measurements is genuinely stochastic. The probability for the mind to jump to a certain state after the measurement is given by Born's rule, on the basis of the *local* (reduced) state of the observer.

d. Once a mind has jumped to a certain state, its successive evolution is ruled by the corresponding component of the state of the observer.

After each measurement the observer's mind chooses only one of the component states of the observer's brain, leaving the others 'uninhabited'. (We shall elaborate on this point in our discussion of the MMI.)

[2]More accurately, rather than of superpositions, one should always talk of 'improper mixtures', since it is a larger system that enters the superposition; but we shall allow ourselves the slips in language.

[3]This view is also advanced by Squires (1990, sections 11.6 and 12.2). See also section 4 below.

Albert and Loewer do not characterise the kind of dualism to which the single mind theory is committed — all they say is that minds 'are not quantum mechanical systems; they are never in superpositions. This is what is meant by saying that they are non-physical' (1988, p. 207).[4]

Albert and Loewer acknowledge that, given the failure of assumption M, a viable Everettian interpretation of quantum mechanics necessarily implies some violation of the supervenience of the mental on the physical.

Nonetheless, they consider the violation of supervenience within the single mind interpretation to be too strong. This arises in two ways:

α. In order to say that minds evolve stochastically, we have to admit transtemporal identity for minds, but the latter is not determined by the physical evolution of the world.

β. Within the single mind view, the mental state does not supervene on the physical state since the same superposed physical state can correspond to different mental states.

While Albert and Loewer consider α to be essential to the solution of the determinism problem, and thus an unavoidable feature of a coherent Everettian theory, they regard β as too high a price to pay, and, in effect, as avoidable: 'on the single mind view, all but one of the elements of a superposition [of brain states] represent, as it were, mindless brains and which element represents a mind is not determined by the physical nature of the underlying brain state and cannot be deduced from the quantum state or from any physical experiment. The non-physicalism of the [single mind view] is especially pernicious. It entails that mental states do not even *supervene* on brain states (or physical states generally) since one cannot tell from the state of a brain what its single mind believes' (Albert and Loewer, 1988, p. 206).

This quotation illustrates the *mindless hulks* problem, on the basis of which Albert and Loewer reject the single mind view. The standard way of presenting the mindless hulks problem in the secondary literature appears

[4] It is difficult to say whether the single mind view implies property or substance dualism. Albert and Loewer seem to refer to a property dualism. However, it could be argued that if the single mind view wants to explain the apparent contradiction that our bodies are typically in superposition states but our minds are always in determined states of beliefs, then perforce it requires *substance* dualism, for a weaker property dualism would not solve the tension between postulates 1 and 2.

to be as follows. If only one component of Alice's brain state is inhabited by her mind, and the same for Bob's brain state, then Alice's mind may be witnessing an uninhabited component of Bob's physical state. In our opinion, however (and, as suggested by the above quotation, also in Albert and Loewer's intentions), the substantial question at the root of the mindless hulks problem is the violation of supervenience entailed by the single mind view, rather than the possibility of interacting with a component of our interlocutor's body without corresponding mind. The latter is an additional (if picturesque!) aspect of the problem, pointed out already by Albert in his book (1992, p. 130). Albert and Loewer feel compelled to bite the bullet of dualism and some non-supervenience of the mental on the physical; however, they do not want to completely give up on the idea of supervenience. The single mind view, in fact, does not only imply the failure of a *metaphysical* supervenience of the mental state on the physical state, as maintained by physicalism, but even of a *nomological* supervenience, which is widely agreed upon even among dualist philosophers.

One should note that Albert and Loewer suggest another way (alternative to the many minds view described below) to restore supervenience. This consists in adopting what they call the 'instantaneous minds' view, which gets rid of the transtemporal identity of minds, i.e. in which there is no matter of fact regarding a unique successor relation between the minds at an earlier and at a later time. If we renounce the transtemporal identity of minds, we have, at each instant, a set of minds that is completely determined by the brain state of the observer, and there is no reason of concern regarding the evolution of minds for the simple fact that there is nothing in the theory that allows us to talk about the evolution of each mind. However, Albert and Loewer reject also this possibility, because this would entail the impossibility of making sense of transition probabilities: 'the cost of surrendering the "trans-temporal identity of minds" would seem to be that we can no longer make sense of statements like "the probability that I will observe spin up on measurement is p" since such statements seem to presuppose that it makes sense to talk of a single mind persisting through time' (Albert and Loewer, 1988, p. 211).

2.3 The Many Minds Interpretation and the mindless hulks problem

The many minds view is thus expressly proposed in order to make up for the 'pernicious non-physicalism' displayed by the single mind view and exemplified by the mindless hulks problem. The single mind view's postulates are still valid, only in place of postulate b we have:

- b′ Every sentient physical system is associated with *an infinity* of minds. If the observer's body is in a superposition state of beliefs, say $|B\rangle = c_1|B_1\rangle + c_2|B_2\rangle + \ldots + c_i|B_i\rangle$, the proportion of minds in state M_k (with $1 < k < i$), corresponding to B_k, is $= |c_k|^2$.

In a nutshell: an uncountable (!) infinity of minds is associated with each brain state and with each measurement the set of minds splits in as many subsets as are the possible results of the measurement. The proportion of minds which ends up in a certain state after the measurement is equal to the squared coefficient of the corresponding brain state. The evolution of minds is still stochastic and governed by Born's rule, applied to each component of the state, thus allowing for the definition of conditional probabilities, say, in the case of successive measurements at times t_1 and t_2 (for instance the conditional probability for a mind seeing up-up at t_2, given that it sees up at t_1).

Note that in order for later memories to faithfully track the earlier observed results, and for the total probabilities to add up to the quantum mechanical ones, not only need the minds possess transtemporal identity, but one must also be able to reidentify the components of the physical state over time (the problem of the transtemporal identity of branches), which in turn presupposes that the components of the state that include the physical correlates of the mental states are subject to decoherence (Bacciagaluppi 2012, Section 3.3). This appears plausible, since they are the seat of memories, or at the very least are correlated with physical records in the environment.[5]

We emphasise that the faithfulness of later records is obviously relevant in the case of Bob's reports to Alice. Should Bob's report not be faithful in the sense above, then there is indeed a sense in which those of Alice's minds that witness Bob's report would be mistaken in attributing (past) mental states to Bob. And this could be taken as a variant of the mindless hulks problem. But if one takes decoherence explicitly into account, we believe this version of the problem is a red herring.

This brings us back to the mindless hulks problem and to the question of why the MMI should be regarded as a solution to the mindless hulks

[5] Decoherence of course is an explicit and crucial component in the many minds or many worlds views by Saunders, Wallace, Zeh and others, to be discussed in the companion paper.

problem.

In the MMI we can distinguish between two different mental states: the *individual* mental state, which is the state of individual minds, and the *total* mental state, i.e. the distribution of mental states among the infinity of the observer's minds.[6] According to Albert and Loewer, while on the one hand the individual state does not supervene on the physical state (for individual minds evolve stochastically and we cannot deduce their state from the physical one), on the other hand the total mental state completely supervenes on the brain state, for the former is uniquely determined by the latter. Thus Albert and Loewer argue that while nomological supervenience of the mental on the physical is not completely restored, the problem is at least downgraded, for there is a sense in which the mental state supervenes on the physical state. In this sense they state: 'we have purchased supervenience of the mental on the physical at the cost of postulating an infinity of minds associated with each sentient being' (1988, p. 207). Barrett (1999 Chap. 7, note 7) stresses that an observer's total mental state does not *strictly* supervene on her physical state. Due to the stochastic dynamics of individual minds, it is in fact always possible (although with probability zero) that the proportion of minds in state M_k corresponding to B_k is different from $|c_k|^2$. This leads Barrett to doubt that the price paid by Albert and Lower is worth the kind of supervenience one gets even in the MMI. Monton (2000) is even more severe and stresses that either the MMI displays supervenience of the mental on the physical, or it does not: *nonstrict* supervenience is no supervenience at all and the non-physicalism of the MMI is therefore just as bad as the one displayed by the single mind view.

Against Monton's conclusion, we still think there is an important sense in which it would be unjust to conclude that the non-physicalism of the MMI is as bad as the one displayed by the single mind view. The issue of how the mental might supervene on the physical draws much of its relevance from what it implies for other relations between mind and body — in particular entailment and explanation. Whether or not to call the relation between mind and body '*quasi* supervenience' risks becoming little more than a terminological issue, if it is not evaluated against such a conceptual background. The fact that in the MMI it is almost always true that the physical state determines the total mental state offers much more scope for entailment and explanation of the mental from the physical than within the

[6] These are called *local* and *global* mental states in (Albert and Loewer, 1988) and (Barrett, 1999). We prefer not to follow this terminology, in order to avoid possible conflation with the issue of locality and non-locality.

single mind view, where 'one cannot tell from the state of a brain what its single mind believes', and, as a result, almost no such entailment and explanation is possible. This should be already sufficient to consider the non-physicalism of the MMI much less pernicious than the one displayed by the single mind view.

3 Locality: Albert and Lower's argument

In the 1988 article where they first propose the MMI, Albert and Loewer make a list of the merits of their theory, last but not least of which is the fact that the many minds view provides an account in which all interactions are *local*. According to Albert, Bell's theorem has no significant consequences for the MMI, for 'Bell proved that there can't be any local way of accounting for the observed correlations between the outcomes of measurements like that; but of course [...] the idea that there ever *are* matters of fact about the "outcomes" of a pair of measurements like that is just what [the MMI] *denies*!' (Albert 1992, p. 132).

Let us take a Bell type experiment. A pair of entangled particles, a and b, in state

$$\frac{1}{\sqrt{2}}(|+\rangle_a|-\rangle_b - |-\rangle_a|+\rangle_b) \tag{1}$$

is sent to two distinct points in space. Then Alice prepares to measure spin-x on a and Bob prepares to measure spin-$(x+\vartheta)$ on b, so that the state of the system composed by Alice, Bob, a and b at the instant just before the measurement is:

$$|\Psi\rangle = |\text{ready}_x\rangle_A\,|\text{ready}_{x+\vartheta}\rangle_B\,\frac{1}{\sqrt{2}}(|+\rangle_a|-\rangle_b - |-\rangle_a|+\rangle_b) \tag{2}$$

where, say, $|\text{ready}_x\rangle_A$ is the state of Alice who is ready to perform a measurement of spin-x. After the two measurements, Alice's and Bob's bodies are entangled with a and b, thus the state of the composite system

a+b+Alice+Bob becomes:

$$
\begin{aligned}
|\Psi'\rangle = \frac{1}{\sqrt{2}}\Big(& \sin\frac{\vartheta}{2}|+_x\rangle_a|+_{x+\vartheta}\rangle_b|+_x\rangle_A|+_{x+\vartheta}\rangle_B \\
& + \cos\frac{\vartheta}{2}|+_x\rangle_a|-_{x+\vartheta}\rangle_b|+_x\rangle_A|-_{x+\vartheta}\rangle_B \\
& - \cos\frac{\vartheta}{2}|-_x\rangle_a|+_{x+\vartheta}\rangle_b|-_x\rangle_A|+_{x+\vartheta}\rangle_B \\
& - \sin\frac{\vartheta}{2}|-_x\rangle_a|-_{x+\vartheta}\rangle_b|-_x\rangle_A|-_{x+\vartheta}\rangle_B\Big)
\end{aligned}
\tag{3}
$$

Suppose Bob performs his measurement first. Half of Bob's minds register + and half register −. But now, the evolution of Alice's minds depends on the *local* (*reduced*) *state of Alice alone*, thus on the evolution of the state $|\mathrm{ready}_x\rangle_A$ to the improper mixture

$$
\rho_A = \frac{1}{2}\left(|+_x\rangle\langle +_x|_A + |-_x\rangle\langle -_x|_A\right) \tag{4}
$$

Therefore, one half of Alice's minds end up in the state 'I am registering +', the other half end up in the state 'I am registering −', *independently of Bob's measurement* (or the outcome registered by any of Bob's minds). The same applies to Bob's minds. At this moment of the experiment there is thus no connection or correlation between Alice's and Bob's minds.

Let us now say that Alice and Bob meet and report their results to each other. At the new instant the state of the composite system has evolved to:

$$
\begin{aligned}
|\Psi''\rangle = \frac{1}{\sqrt{2}}\Big(& \sin\frac{\vartheta}{2}|+_x\rangle_a|+_{x+\vartheta}\rangle_b| +_x +'\rangle_A| +_{x+\vartheta} +'\rangle_B \\
& + \cos\frac{\vartheta}{2}|+_x\rangle_a|-_{x+\vartheta}\rangle_b| +_x -'\rangle_A| -_{x+\vartheta} +'\rangle_B \\
& - \cos\frac{\vartheta}{2}|-_x\rangle_a|+_{x+\vartheta}\rangle_b| -_x +'\rangle_A| +_{x+\vartheta} -'\rangle_B \\
& - \sin\frac{\vartheta}{2}|-_x\rangle_a|-_{x+\vartheta}\rangle_b| -_x -'\rangle_A| -_{x+\vartheta} -'\rangle_B\Big)
\end{aligned}
\tag{5}
$$

where, say, the state $| +_x +'\rangle_A$ is Alice's brain state that registers both + on a and Bob's + report.

Here, after the interaction with Alice, $\cos^2(\vartheta/2)$ of Bob's minds believe that her result is the opposite of Bob's result. The probability for each of Bob's minds to believe that the two results are opposite is $\cos^2(\vartheta/2)$, and the same is true of Alice's minds. Furthermore, each of Bob's minds that has registered + has the conditional probability $\cos^2(\vartheta/2)$ of witnessing Alice's body reporting a $-$ result. This is because the $|+_{x+\vartheta}\rangle$ and $|-_{x+\vartheta}\rangle$ components of Bob's state have decohered, and it makes sense to talk about the $|+_{x+\vartheta}\rangle$ component of Bob's state splitting further into $|+_{x+\vartheta}+'\rangle$ and $|+_{x+\vartheta}-'\rangle$. Note that this is due to the *local* evolution of Bob's state when Bob meets Alice. Only at this point can we talk of matching up subsets of Alice's minds with subsets of Bob's minds (Albert and Loewer 1988, p. 210; Bacciagaluppi 2002, p. 111).

The two sets correspond to each other, in that the mental states in these sets encode appropriately matching results. The corresponding sets have the same measure, so there exist ways of pairing off Alice minds with Bob minds in a measure-preserving way, but no particular way of doing so has any special significance. In particular, it is immaterial for the purposes of the MMI which of Bob's + minds witness Alice reporting a + result and which witness Alice reporting a $-$ result. Note again that these 'correspondences' in the Bell-type setting are reproduced by means of the local evolution of the brain states plus a completely local dynamics for the minds. We shall return to these corresponding sets below.

4 The mindless hulks problem and nonlocality

We now turn to how the interpretation of the mindless hulks problem, and how the MMI is supposed to deal with it, has ramifications for the question of the local nature of the MMI.

We have seen that, if the mindless hulks problem is interpreted as a problem of lack of supervenience and physicalism, as we think it should, the solution provided by the MMI is entirely local.[7] Often, however, the mindless hulks problem is stated as being that every time I talk with a human being, there is the possibility that what I am interacting with is not a sentient being, but a 'mindless hulk'.

Now, if this were all there was to the mindless hulks problem, there would be a possible solution *already within the single mind view*, and that

[7]As we have seen, it is debatable whether the kind of supervenience purchased in the MMI is worth the cost of its multiplication of minds. Here we consider only how different understandings of the problem can lead to different results even with respect to the issue of locality.

is to admit *correlations* between observers' minds. Let us take the EPR experiment of section 3, and let us say that Alice and Bob measure the spin in the same direction. If we allow the evolution of Alice's mind to depend on whether Bob's mind has registered $+$ or $-$, then we can guarantee that the results that Alice's and Bob's minds register are always opposite. In this way the mindless hulks problem in the above sense is solved. Note that this kind of correlations might lead to a problem with nonlocality.[8] Indeed, the single mind interpretation is subject to Bell's theorem, since there are unique measurement outcomes, at least in the observers' minds. However, note that the unconditional probability for Alice's mind registering either result is $1/2$, the same as if Bob had performed no measurement (or a different measurement). There are thus nonlocal correlations between Bob's mind and Alice's mind, but no dependence of Alice's mind on Bob's settings (or vice versa). The resulting nonlocality consists only in a violation of outcome independence, but not of parameter independence, exactly as in standard quantum mechanics.

However, this alternative does not solve the mindless hulks problem *as a problem of nonphysicalism*, and indeed it is never considered by Albert and Loewer in their 'step by step' construction of the MMI. This alternative would make the problem of supervenience even worse, since in the single mind view a correlation between minds could not be 'mediated' or supported by the physical state, for minds do not supervene on it. But if the physical state is not able to support these correlations, a reference to a *direct connection between minds* is necessary in order to account for them, and this would imply a definitely bizarre kind of dualism.

If, after all, one is worried by the more picturesque aspect of the mindless hulks problem, one might think that the MMI as described above does not provide a complete solution of the problem. The worry is presumably that the mere *possibility* of pairings-off of Alice's and Bob's minds witnessing the same pairs of results is not substantial enough, and that there should be some *actual* pairing-off of the minds to ensure that each of Alice's minds is communicating with an actual mind behind what it perceives as Bob's body.

[8] This is effectively the position suggested by Squires (1990, sections 11.6 and 12.2), who also speculates about whether different observers' minds might be best understood as individual minds or as parts of a single universal consciousness. In this sense the 'nonlocality' of the correlations between the minds might be alleviated. One might not be bothered by this nonlocality also if one adopts a mental monist position (e.g. Monton 2000, 2007).

Under this reading, the MMI as a solution of the mindless hulks problem might indeed have to include some kind of correlations between Bob's minds and Alice's minds. Specifically, one could introduce correlations between pairs of Alice's and Bob's individual minds, as in the above variant of the single mind view. In this case, each of Alice's minds would 'communicate' just with one of Bob's minds, and their evolutions would depend on each other as in the single mind view. This would take care of both readings of the mindless hulks problem. However, it would do so by giving up the locality of Albert and Loewer's original MMI and by embracing what we have described as a bizarre kind of direct dependence between the minds.

A variant of this worry appears also in Hemmo and Pitowsky (2003). Hemmo and Pitowsky reject the above idea of correlations, because of the problems with locality, but suggest that in order to solve the mindless hulks problem, one has to introduce some weaker form of correlations, claiming that even these 'weak minds-correlations' require a price to be paid in terms of a weak form of nonlocality (to which we shall return in section 5).

We suggest, however, that even on the 'picturesque' reading, the mindless hulks problem does not require the introduction of correlations over and above the correspondence of the subset of Bob minds that witness a particular Alice report and the subset of Alice minds that have witnessed the corresponding result (and vice versa). One only has to note that the connection between the 'corresponding' minds is mediated by their physical state. Specifically, each of the components of the state (5) gives rise both to Alice's experience and report of her own result, and to Bob's witnessing of that report. As regards 'hulks', the situation is no weirder than if Alice and Bob had each measured an electron that was in an eigenstate of spin (where there also is no communication between *individual* minds, but each mind has *all* of the other's minds as interlocutors).

Note that if Alice and Bob meet and compare results, then the correspondence established at this point can be read backwards to the time when Alice and Bob originally performed their measurements: there will be subsets A_k of Alice minds and B_k of Bob minds that will 'follow the same branch' in the future. (If one will, this is a 'many worlds' reading of the MMI: Alice's measurement not only splits her brain in two components, but the whole world, including Bob's brain. See also the companion paper.) However, this allows us *neither* to specify in advance which subset B_k will follow the same branch as A_k in the future, *nor* to say that if Alice and Bob had performed different measurements than those actually performed, a different subset B'_k would have followed the same branch as A_k. Given the stochastic dynamics of minds, in fact, *which* minds follow a determinate branch is a contingent fact, and the same is valid for *which* minds will accompany A_k in the future and *which* minds would have accompanied A_k in a counterfactual situation. In the next section we shall analyse this claim in more detail.

5 Hemmo and Pitowsky's weak nonlocality

Our discussion above endorses Albert and Loewer's claim that their version of the MMI does not run into trouble with Bell's theorem, specifically that it exhibits neither violations of outcome independence (unlike the single mind view) nor violations of parameter independence (indeed, the measure of the set of Alice's minds that witness a certain outcome of her measurement does not depend on the settings chosen by Bob for his measurement). There is, however, another notion of nonlocality that does not apply to the cases covered by Bell's theorem, but that might be exhibited by the MMI. This is Hemmo and Pitowsky's (2003) notion of *weak nonlocality*, namely the idea that *which* of Alice's minds witness a particular outcome in general *does* depend on Bob's settings. (They call this 'weak nonlocality', because, unlike violations of parameter independence, it does not lead to superluminal signalling.) Hemmo and Pitowsky claim that if one assumes 'weak minds-correlations' (in order to solve the mindless hulks problem), then the MMI will exhibit weak nonlocality. We shall not go into the details of whether Hemmo and Pitowsky's claim is correct.[9] We do wish, however, to discuss in more detail whether anything like weak nonlocality might arise in the picture(s) of the MMI we have been discussing above, with particular reference not only to the EPR case, but also to the case that is singled out as particularly interesting by Hemmo and Pitowsky, namely the case of the so-called GHZ state (after Greenberger, Horne and Zeilinger, 1989).

We first consider again explicitly the Bell-type situation of section 3. We have already argued there that the evolution of the minds is indeed completely local, while recovering at least the illusion of an interaction between the corresponding minds. Our result does not change if we ask explicitly also which minds follow which branch.

[9] We believe not, but we are not exactly clear on what Hemmo and Pitowsky's notion of 'weak minds-correlations' is in the first place.

For instance, it is true that the set of Alice's minds that (say) witness both a $+$ result in Alice's experiment and Bob reporting a $-$ result will in general depend on the setting $x+\vartheta$ chosen by Bob (indeed, the measure of this set will depend on it). However, this is the set of Alice's minds that will witness a $-$ report *only* when Alice meets up with Bob *in the future*, and there is no physical matter of fact about this set before that time, and so no nonlocal influence of Bob's choice on this partition of Alice's minds.

Moreover, there is also not necessarily any influence of Bob's choice on which of Alice's minds witnesses a $+$ report in Alice's experiment, since, as long as later $\cos^2(\vartheta/2)$ of both Alice's $+$ minds and Alice's $-$ minds witness an opposite report from Bob, *which* of Alice's minds are going to witness $+$ or $-$ is immaterial (and similarly for Bob's minds). Explicitly, one could even imagine that the *same* half of Alice's minds always registers a $+$ result independently not only of Bob's but even of Alice's own settings.[10] Indeed, one might argue that the closest possible worlds to that with the actual settings are those in which the same minds witness $+$ or $-$ as in the actual world, even with a different setting.

Thus, in the standard EPR case weak nonlocality does not follow. The main example of weak nonlocality in Hemmo and Pitowsky, however, is the GHZ case; so we turn to an explicit analysis also of this case.

Consider the GHZ state:

$$\frac{1}{\sqrt{2}}(|{+_z}\rangle_1|{+_z}\rangle_2|{+_z}\rangle_3 + |{-_z}\rangle_1|{-_z}\rangle_2|{-_z}\rangle_3) \tag{6}$$

and let us say that three observers, Alice, Bob and Carol, respectively measure the spin of electrons 1, 2 and 3 in the same direction x. The quantum mechanical prediction is that the product of the spin of the three electrons in the direction x is always equal to -1. Always according to quantum mechanics, the state of the electrons and observers after the three measurements is:

$$\begin{aligned}|\Psi\rangle = \frac{1}{2}\Big(&|{-_x}\rangle_1|{+_x}\rangle_2|{+_x}\rangle_3|{-_x}\rangle_A|{+_x}\rangle_B|{+_x}\rangle_C \\ &+|{+_x}\rangle_1|{-_x}\rangle_2|{+_x}\rangle_3|{+_x}\rangle_A|{-_x}\rangle_B|{+_x}\rangle_C \\ &+|{+_x}\rangle_1|{+_x}\rangle_2|{-_x}\rangle_3|{+_x}\rangle_A|{+_x}\rangle_B|{-_x}\rangle_C \\ &+|{-_x}\rangle_1|{-_x}\rangle_2|{-_x}\rangle_3|{-_x}\rangle_A|{-_x}\rangle_B|{-_x}\rangle_C\Big)\end{aligned} \tag{7}$$

[10] In order to formulate this unambiguously, of course, we have to stipulate for each pair of opposite directions which is to count as *the* direction the spin is measured along (no way of doing this will be natural, but this is beside the point).

Let us call this case *scenario 1*, and the cases in which Alice, Bob and Carol measure the spin either in the directions (x, y, y) or (y, x, y) or(y, y, x), respectively, *scenarios 2*, *3* and *4*. In these other cases, the product of the results of the measurements is equal to 1.

According to Hemmo and Pitowsky, assuming their weak minds-correlations hold, after the measurements of scenario 1 the set $M_A \times M_B \times M_C$ of triples of minds of Alice, Bob and Carol is partitioned into four subsets of triples:

$$\begin{aligned} M_A \times M_B \times M_C = \Big(& M_A^-(1) \times M_B^+(1) \times M_C^+(1) \cup \\ & M_A^+(1) \times M_B^-(1) \times M_C^+(1) \cup \\ & M_A^+(1) \times M_B^+(1) \times M_C^-(1) \cup \\ & M_A^-(1) \times M_B^-(1) \times M_C^-(1) \Big) \end{aligned} \tag{8}$$

where, say, $M_A^+(1)$ represents the set of Alice's minds that register + in scenario 1. Always according to Hemmo and Pitowsky, analogous partitions arise in scenarios 2, 3 and 4. (Note that without some such supplementary assumption, the supposed partition (8) does not exhaust $M_A \times M_B \times M_C$ as a set!)

Hemmo and Pitowsky now consider the common sub-partition of the above four partitions that consists of sets of the form:

$$\begin{aligned} & M_A^-(1) \times M_B^+(1) \times M_C^+(1) \cap \\ & M_A^-(2) \times M_B^-(2) \times M_C^+(2) \cap \\ & M_A^+(3) \times M_B^-(3) \times M_C^-(3) \cap \\ & M_A^+(4) \times M_B^+(4) \times M_C^+(4) \end{aligned} \tag{9}$$

etc., i.e. intersections of the sets of triples of minds that follow a specific branch in each scenario.

There exist $4^4 = 256$ possible combinations of this kind, in each of which there is an observer performing the same spin measurement in two different scenarios but obtaining different results (in the example of (9), it is Alice in scenarios 3 and 4).

Now, according to Hemmo and Pitowsky, the 256 elements 'exhaust every logical possibility. Therefore, at least one of those sets has probability $\geqslant \frac{1}{256}$' (p. 241).[11] If this is correct, then it is necessarily the case that a non-zero measure set of minds of at least one observer registers *different*

[11]At first sight, one might wonder about what probability measure is being used here,

measurement results (for the same measurement), depending on the *settings* chosen by the other observers, which would indeed be an example of Hemmo and Pitowsky's weak nonlocality.

At least in the readings of the MMI we have considered above, however, the conclusion does not follow.

For instance, consider the case in which individual minds are correlated (which is not Albert and Loewer's understanding of the MMI). That is, we consider the quantum state as defining probabilities for the evolution of the triples of minds, rather than just for each observer's minds separately. In this case, (8) is indeed a partition of all the triples that are given non-zero probability by the GHZ state if the measurements of scenario 1 are carried out (and similarly for scenarios 2 3 and 4). But the GHZ state defines these probabilities contextually for each scenario separately, and in this reading of the MMI there is no joint probability distribution on events of the type (9). (Non-existence of such joint probabilities is discussed in the celebrated book by Pitowsky himself (1989).) So, although in this case we have nonlocality in the sense of outcome dependence, no weak nonlocality follows.

Now consider our corresponding sets (i.e. what we take to be Albert and Loewer's original reading of the MMI). Say, in scenario 1, consider Alice's minds registering $-$ and eventually witnessing Bob's and Carol's minds registering $+$ (call this set $M_a^{-++}(1)$), and match these minds with Bob's minds and Carol's minds that also register or eventually witness the same triple of results. Then, the sets of the form $M_a^{-++}(1) \times M_b^{-++}(1) \times M_c^{-++}(1)$, etc. (for all four allowable triples), do not form a partition of $M_a \times M_b \times M_c$, and the argument breaks down at the first step. Indeed, the only requirement on the corresponding sets is that, say, in scenario 1 Alice's $+$ minds eventually divide evenly (with probability 1/2, that is) into $+-+$ minds and $++-$ minds, etc. This requirement is satisfiable irrespectively of which of Alice's minds originally register $+$. (And the further evolution takes place only when Alice meets Bob and/or Carol, or otherwise learns about their results, and is thus also to be thought of as local.) Indeed, we can pick at random one quarter of Alice's minds, and stipulate arbitrarily not only that these are the minds that register or witness $+-+$ in scenario 1, but also that they register any allowable triple of results in any of the other three scenarios (and similarly with the rest of Alice's minds and with Bob's

given that these sets are defined with reference to incompatible measurement scenarios. But if the sets of the form 9 indeed should form a partition of $M_a \times M_b \times M_c$, any probability measure on $M_a \times M_b \times M_c$ for which such sets are measurable will do.

and Carol's minds). As long as each of these sets have the same size, the matching requirement can be explicitly met, and the probabilistic evolution of the minds in each individual scenario is as postulated in the MMI.

We can indeed see that the assumptions of the MMI are independent of weak locality, i.e. we can explicitly construct models that are free from weak nonlocality as well as models that exhibit it.

For instance, one can stipulate that the *same* set of Alice's (and Bob's and Carol's) minds register $+$ (or $-$) in *all* of the above scenarios.[12] In this case, all of the sets of the form (9) are empty, and no mind flips its sign depending on the measurement performed by other observers, i.e. there is no weak nonlocality.

Or one could equally stipulate that the *same minds be matched in all scenarios* above. Thus, for instance, the intersection (9) would turn out to be just the set $M_a^{-++}(1) \times M_b^{-++}(1) \times M_c^{-++}(1)$, and all of Alice's $--+$ minds in scenario 3 flip when we move to scenario 4. In this case we do have weak nonlocality (even though the sets in (9) do not form a partition of $M_a \times M_b \times M_c$, nor *a fortiori* the sets in (8)).

Note that any of these stipulations are entirely formulated at the level of counterfactuals, and it is for this reason that they are independent of the postulates of the MMI, which are formulated at the level of each individual scenario.

The assumption that the same minds be matched in all scenarios is thus compatible with Albert and Loewer's MMI, and could be added to it, perhaps in the more general (and to be made precise) form that as many minds as possible that are matched in the actual world should be matched in counterfactual situations. This could be read as a requirement on the closeness of possible worlds (but so is the requirement that the results witnessed by individual minds should not flip). Or its violation could be read as a distant cousin of the mindless hulks problem in its more picturesque reading: Alice's minds will interact with different sets of minds depending on which scenario is chosen for the measurements. But we doubt this would cause Alice's minds very great existential pangs.[13]

[12] And even that the same subset of Alice's $+$ minds further witness Bob registering $+$ in each scenario (it is then fixed by the respective scenario what this further subset of Alice's minds will witness Carol as registering).

[13] Note that we are not suggesting that this condition (which as mentioned also does not turn the sets (9) into a partition of $M_a \times M_b \times M_c$) correctly represents Hemmo and Pitowsky's notion of 'weak minds-correlations'.

As a final remark: even on the assumption that the same minds are matched in all scenarios, one might wonder how to characterise the dependence of the partition of the minds on the settings chosen by the other observers, i.e. whether it is after all a form of nonlocality. Indeed, one should take into account also the cases in which none of the observers have met or otherwise learned of one another's results. In this case, one might argue that the assumption is vacuous. At the very least, one might argue that the constraint comes in place only at a later time, when (and if) the observers learn of one another's results, so that the most natural interpretation of the constraint may be not in terms of action-at-a-distance (spacelike) but of backwards causation (timelike), thus arguably more compatible with relativity. However, we shall not press this point.

References

Albert, David. 1992. *Quantum Mechanics and Experience*. Cambridge-London, Harvard University Press.

Albert, David and Barry Loewer. 1988. Interpreting the Many Worlds Interpretations. *Synthese* LXXXII (2): 195–213.

Atmanspacher, Harald. 2011. Quantum Approaches to Consciousness. In *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), Edward N. Zalta (ed.), `http://plato.stanford.edu/archives/sum2011/entries/qt-consciousness/`.

Bacciagaluppi, Guido. 2002. Remarks on Space-time and Locality in Everett's Interpretation. In *Non-Locality and Modality*, eds. Tomasz Placek and Jeremy Butterfield, Dordrecht, Kluwer, 105–122.

Bacciagaluppi, Guido. 2012. The Role of Decoherence in Quantum Mechanics. In *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.), `http://plato.stanford.edu/archives/win2012/entries/qm-decoherence/`.

Barrett, Jeffrey A. 1999. *The Quantum Mechanics of Minds and Worlds*. Oxford, Oxford University Press.

Barrett, Jeffrey A. 2006. A Quantum-Mechanical Argument for Mind-Body Dualism. *Erkenntnis* 65: 97–115.

DeWitt, Bryce S. 1971. The Many-Universes Interpretation of Quantum Mechanics. In *Proceedings of the International School of Physics 'Enrico*

Fermi', Course IL: Foundations of Quantum Mechanics, ed. Bernard d'Espagnat. New York, Academic Press. Reprinted in DeWitt and Graham (1973), 167–218.

DeWitt, Bryce S., and Neill Graham, eds. 1973. *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton, Princeton University Press.

Everett, Hugh III. 1957. 'Relative State' Formulation of Quantum Mechanics. *Reviews of Modern Physics.* XXIX: 454–462.

Everett, Hugh III. 1973. The Theory of the Universal Wave Function. In DeWitt and Graham (1973), 1–140.

Greenberger, Daniel, Michael Horne, and Anton Zeilinger. 1989. Going beyond Bell's theorem. In *Bell's theorem, quantum theory, and conception of universe*, ed. Menas Kafatos. Dordrecht, Kluwer, 73–76.

Hemmo, Meir and Itamar Pitowsky. 2003. Probability and Nonlocality in Many Minds Interpretations of Quantum Mechanics. *British Journal for the Philosophy of Science* 54(2): 225–243.

Monton, Bradley. 2000. On Dualistic Interpretations of Quantum Mechanics, `http://philsci-archive.pitt.edu/2913/`.

Monton, Bradley. 2007. Common-Sense Realism and the Unimaginable Otherness of Science. *Principia* 11(2): 117–126.

Pitowsky, Itamar. 1989. *Quantum Probability, Quantum Logic.* Lecture Notes in Physics, vol. 321. Berlin, Springer.

Saunders, Simon. 1995. Time, Quantum Mechanics and Decoherence. *Synthese* 102(2): 235–266.

Squires, Euan. 1990. *Conscious Mind in the Physical World.* Bristol, Adam Hilger.

Wallace, David. 2002. Worlds in the Everett Interpretation. *Studies in History and Philosophy of Modern Physics.* 33(4): 637–661.

Wallace, David. 2003. Everett and Structure. *Studies in History and Philosophy of Modern Physics.* 34(1): 87–105.

Wallace, David 2012. *The emergent multiverse.* Oxford, Oxford University Press.

Wigner, Eugene P. 1961. Remarks on the Mind-Body Question, in *The Scientist Speculates*, Good, I. J. (ed.). London, Heinemann, 284–320. Reprinted in J. A. Wheeler and W. H. Zurek, eds., *Quantum Theory and Measurement.*

Princeton, Princeton University Press. 1983, 168–181.

Zeh, Hans-Dieter. 2000. The Problem of Conscious Observation in Quantum Mechanical Description. *Foundations of Physics Letters* 13(3): 221–233.